Woods Hole Oceanographic Institution

Biology Department Seminar

Thursday, October 26, 2023 – 12:00 Noon

Metapathpredict: A Machine Learning-Based Tool for Prediction of Metabolic Modules in Incomplete Bacterial Genomes

David Geller-McGrath

WHOI-MIT Joint Program Student

Microorganisms play a key role in all major biogeochemical cycles on Earth. Accurate and more complete identification of microbial metabolic pathways within genomic data is crucial to understanding their potential activities. Together with assessment of gene expression, it is increasingly possible to gain important insight into feedbacks between microorganisms and the chemistry of their environment as well as into their interactions with other organisms. Reconstruction of complete microbial metabolic pathways using 'omics data from environmental samples remains challenging due to the incomplete nature of most genomic data for individual taxa. Computational pipelines for pathway reconstruction that utilize machine learning methods to predict the presence or absence of KEGG modules in incomplete genomes are lacking. MetaPathPredict is a new software tool that incorporates machine learning models to predict the presence of complete KEGG modules within bacterial genomes using gene annotation data and information from the KEGG module database. Benchmarks show that MetaPathPredict makes robust predictions of KEGG module presence in highly incomplete bacterial genomes. MetaPathPredict has also been applied to MAGs from poorly characterized candidate phyla recovered from the Guaymas Basin deep subsurface. This new tool facilitates more complete and accurate reconstruction of the metabolic potential of bacterial genomes, which can inform analysis of environmental genomic datasets.

HYBRID! **In person**: Clark 201 **Zoom:** https://whoi-edu.zoom.us/j/99399604868 Meeting ID: 993 9960 4868 **By phone**: Find your local number: https://whoi-edu.zoom.us/u/aeH38NVF4H